Natural Language Processing and Word Embeddings

Prof. Giuseppe Serra

University of Udine

Document analysis

Named Entities Extraction

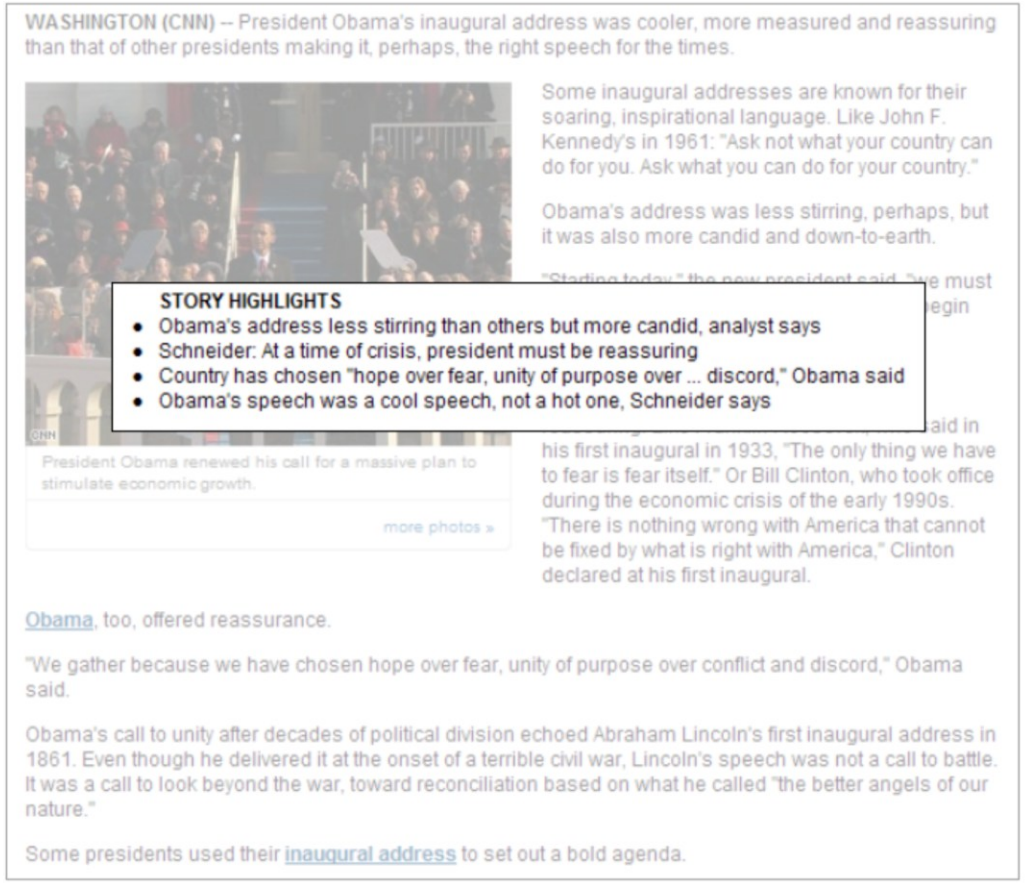
Apple CEO TIM Cook introduces 2 new large iPhones, Smart Watch at Cupertino

Organizzation

Person

Location

Summarization



WASHINGTON (CNN) -- President Obama's inaugural address was cooler, more measured and reassuring than that of other presidents making it, perhaps, the right speech for the times.

Some inaugural addresses are known for their soaring, inspirational language. Like John F. Kennedy's in 1961: "Ask not what your country can do for you. Ask what you can do for your country."

Obama's address was less stirring, perhaps, but it was also more candid and down-to-earth.

STORY HIGHLIGHTS

- Obama's address less stirring than others but more candid, analyst says
- Schneider: At a time of crisis, president must be reassuring
- Country has chosen "hope over fear, unity of purpose over ... discord," Obama said
- Obama's speech was a cool speech, not a hot one, Schneider says

President Obama renewed his call for a massive plan to stimulate economic growth.

more photos »

his first inaugural in 1933, "The only thing we have to fear is fear itself." Or Bill Clinton, who took office during the economic crisis of the early 1990s. "There is nothing wrong with America that cannot be fixed by what is right with America," Clinton declared at his first inaugural.

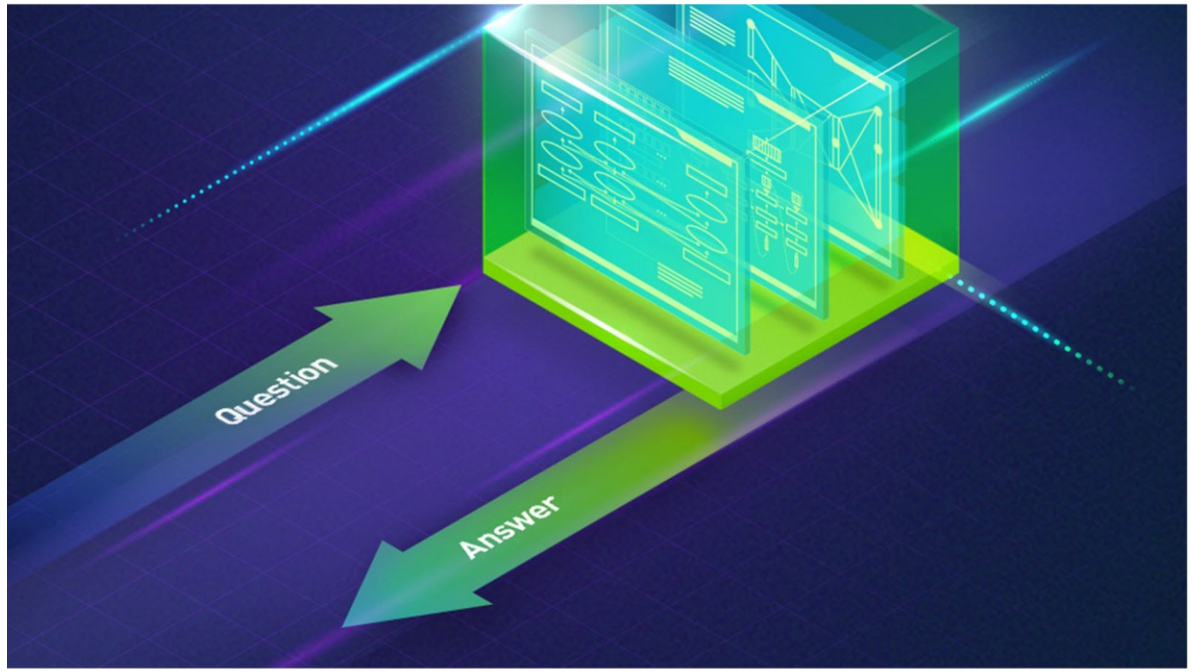
Obama, too, offered reassurance.

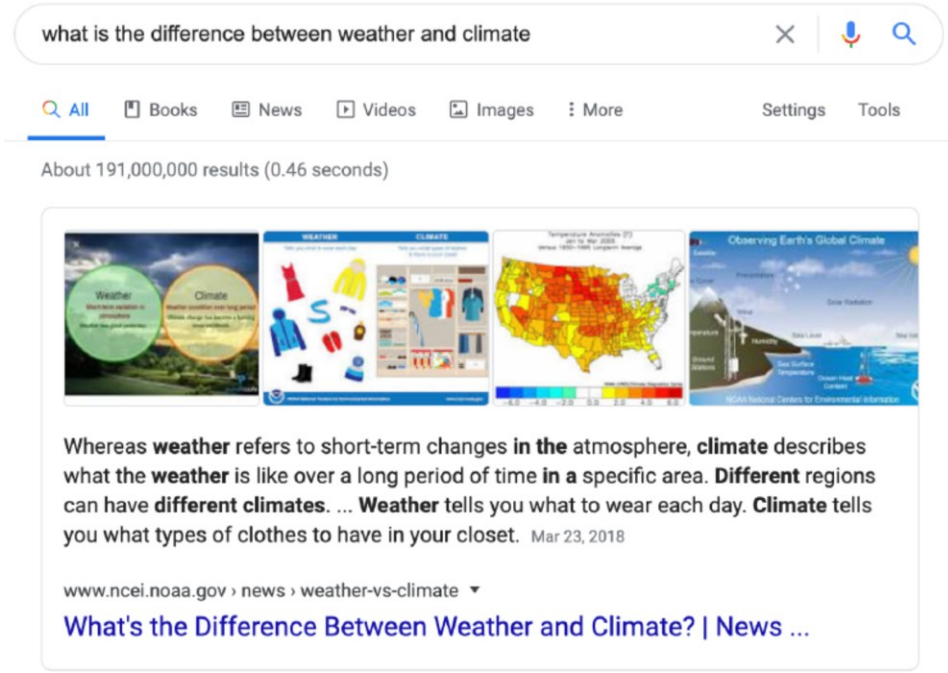
"We gather because we have chosen hope over fear, unity of purpose over conflict and discord," Obama said.

Obama's call to unity after decades of political division echoed Abraham Lincoln's first inaugural address in 1861. Even though he delivered it at the onset of a terrible civil war, Lincoln's speech was not a call to battle. It was a call to look beyond the war, toward reconciliation based on what he called "the better angels of our nature."

Some presidents used their inaugural address to set out a bold agenda.

Question Answering





Social Media Intelligence

forbes.com

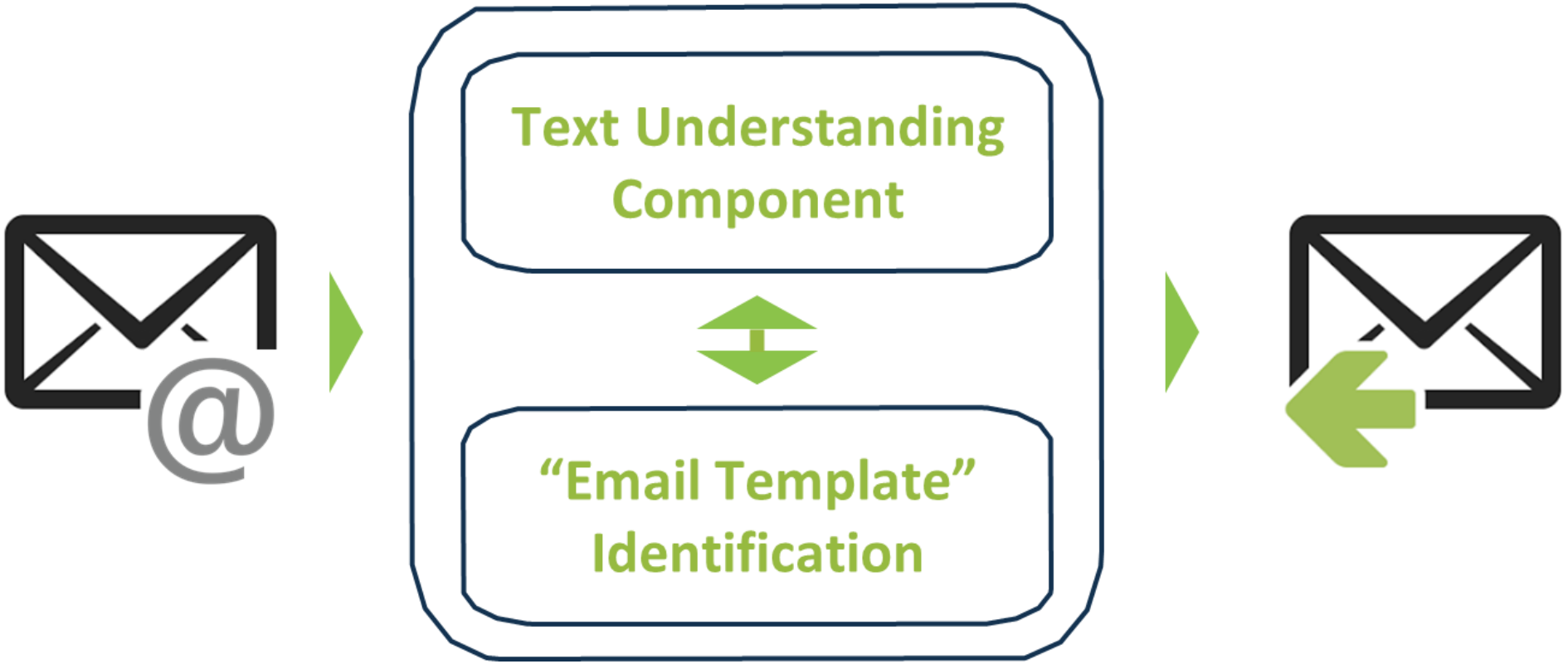
Chiara Ferragni, Instagram's It Girl

Since then, Ferragni and The Blonde Salad have gone from strength to strength: she's launched her own line of apparel and accessories, and employs 20 staffers for a fashion site that competes with traditional media platforms.

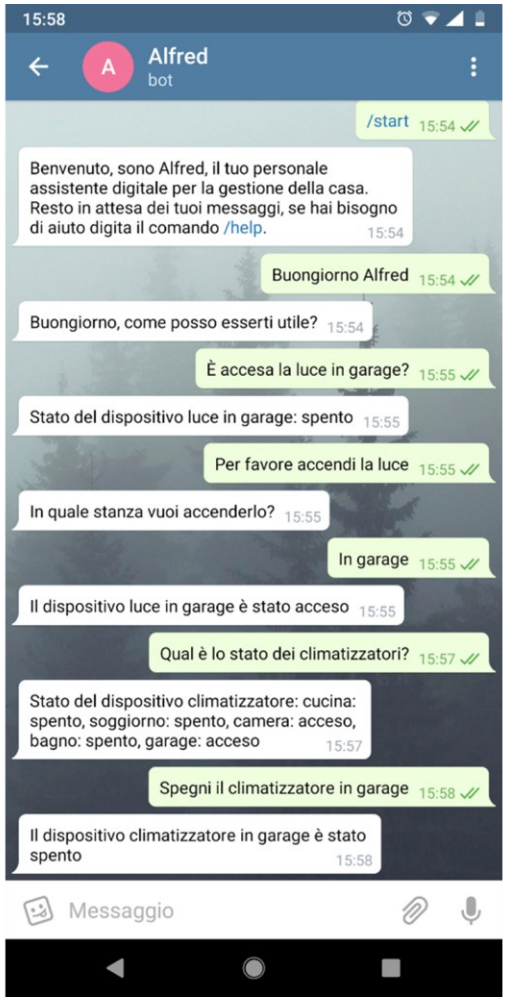
The 30-year-old lands at #1 on Forbes' Top Influencers list in the Fashion category, joining other impressive

Video Machine Translation

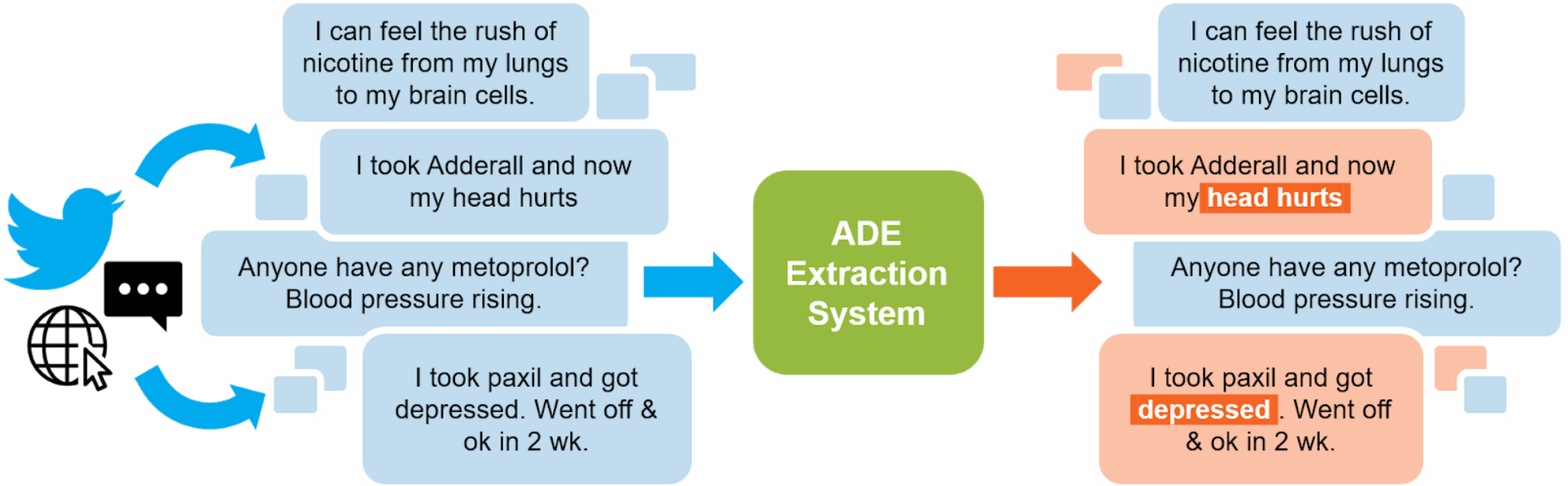
Email Auto Responder (AILAB-Udine)



ChatBot

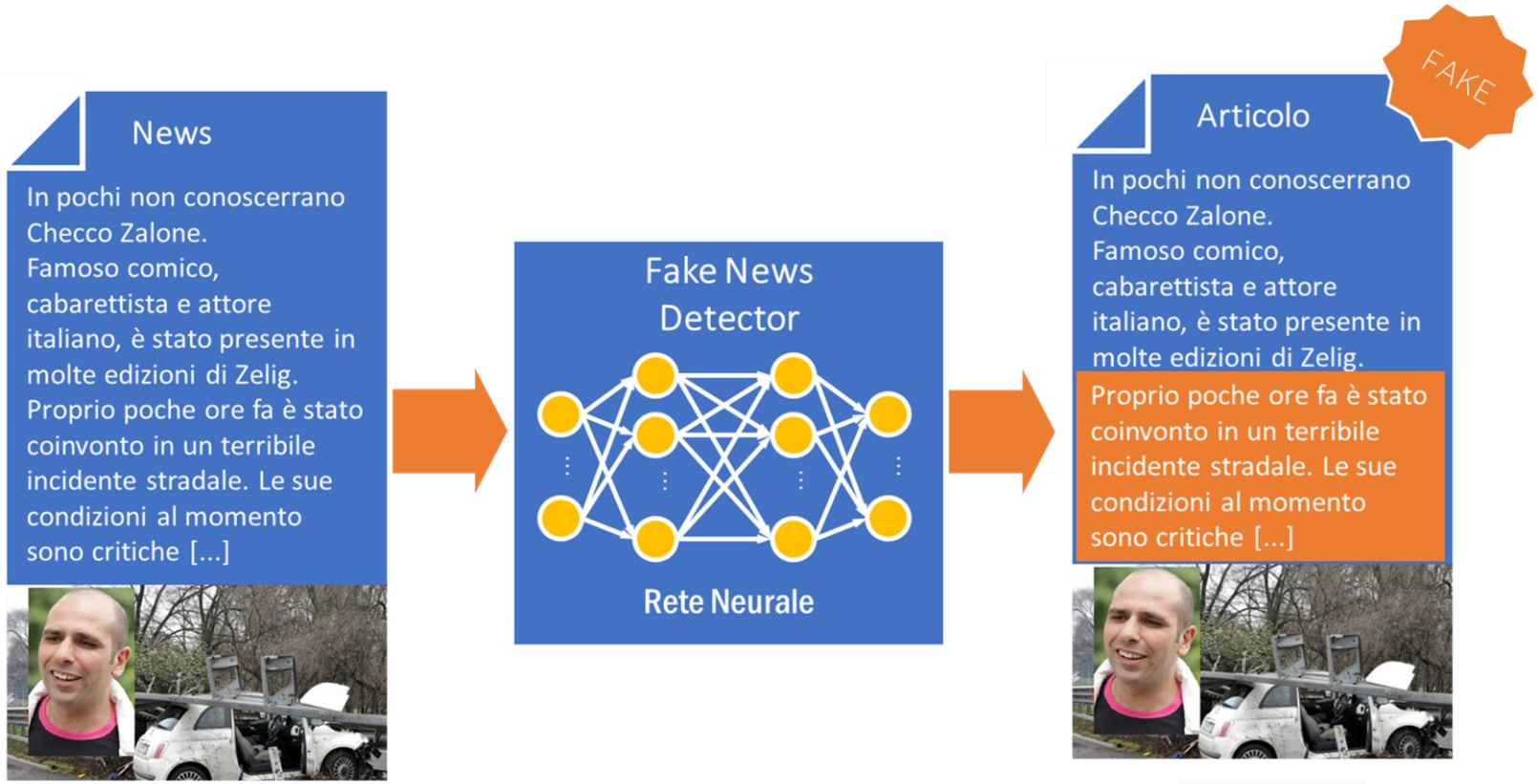

15:58
Alfred
bot
/start 15:54
Benvenuto, sono Alfred, il tuo personale assistente digitale per la gestione della casa. Resto in attesa dei tuoi messaggi, se hai bisogno di aiuto digita il comando /help. 15:54
Buongiorno Alfred 15:54
Buongiorno, come posso esserti utile? 15:54
È accesa la luce in garage? 15:55
Stato del dispositivo luce in garage: spento 15:55
Per favore accendi la luce 15:55
In quale stanza vuoi accenderlo? 15:55
In garage 15:55
Il dispositivo luce in garage è stato acceso 15:55
Qual è lo stato dei climatizzatori? 15:57
Stato del dispositivo climatizzatore: cucina: spento, soggiorno: spento, camera: acceso, bagno: spento, garage: acceso 15:57
Spegni il climatizzatore in garage 15:58
Il dispositivo climatizzatore in garage è stato spento 15:58
Messaggio

COVID-19 Vaccine (AILAB-Udine)

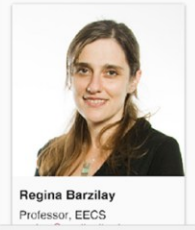


http://ailab.uniud.it/covid-vaccines/

Fake News (AILAB-Udine)



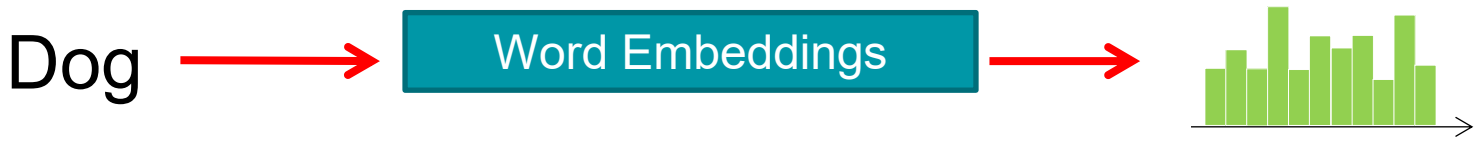
Project in collaboration with Massachusetts Institute of Technology (MIT) – Boston USA



Word Embedding

Word Representation

Word representation techniques represent words as feature vectors.



Combined with Deep Learning these techniques are state of the art in NLP in almost all the tasks.

In the following slides we will see basic and more advance techniques.

Basic Word Representation - Integers

The **simplest way to represent words as numbers** is for a given **vocabulary to assign a unique integer to each word**.

The **set of unique words used in the text corpus** is referred to as the **vocabulary.**

Word	Number
a	1
able	2
about	3
...	...
hand	615
...	...
happy	621
...	...
zebra	1000

Basic Word Representation - Integers

This numbering scheme is simple,

One of the problems is that the order of the words,

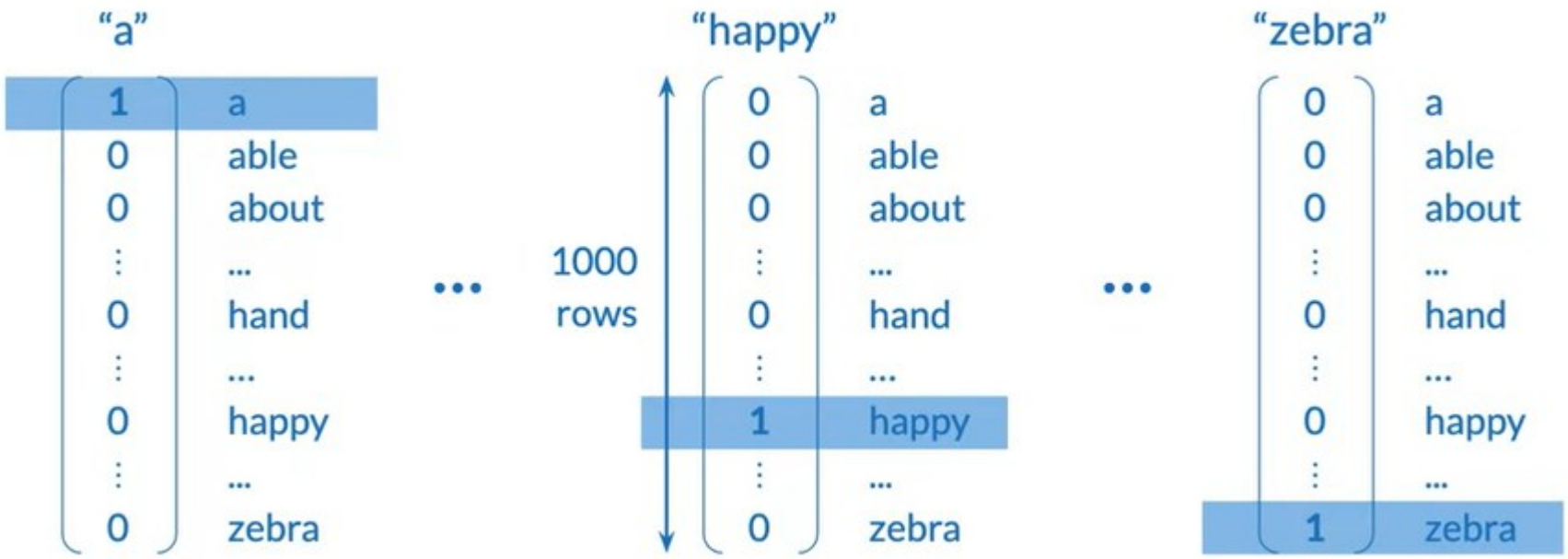
For instance, **alphabetical order doesn't make much sense from a semantic perspective**.

\+ Simple

\- Ordering: little semantic sense

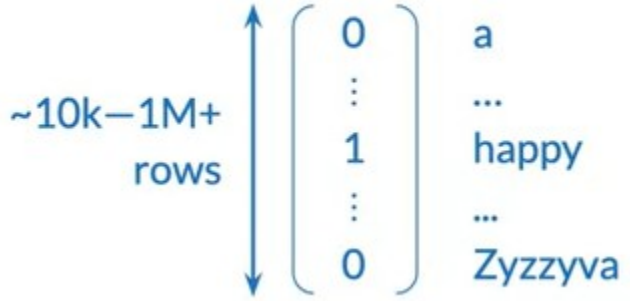
Basic Word Representation – One-hot Vectors

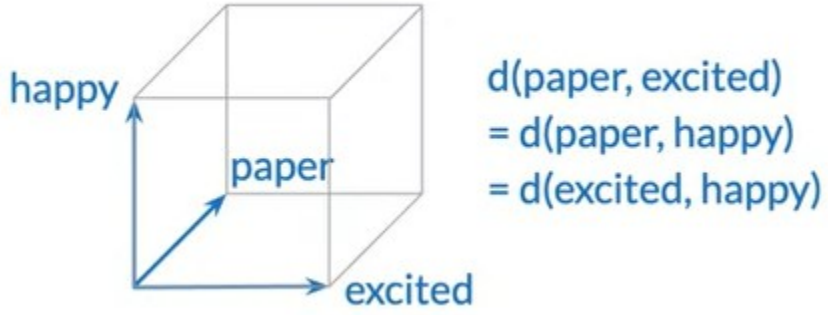
One-hot Vectors represent the words using a column vector where each element corresponds to a word of the vocabulary.



Basic Word Representation – One-hot Vectors

+ Simple
+ No implied ordering
- Huge vectors
- No embedded meaning



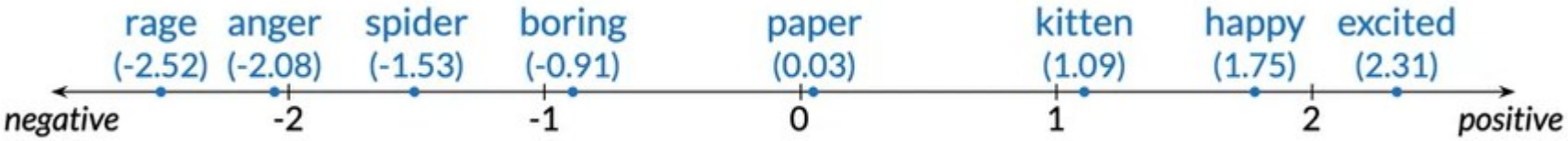


Basic Word Representation – Word Embedding

Word embeddings are vectors that are carrying meaning.

Consider a horizontal number line like this. Words on the left are considered negative in some way and words on the right are considered positive in some.

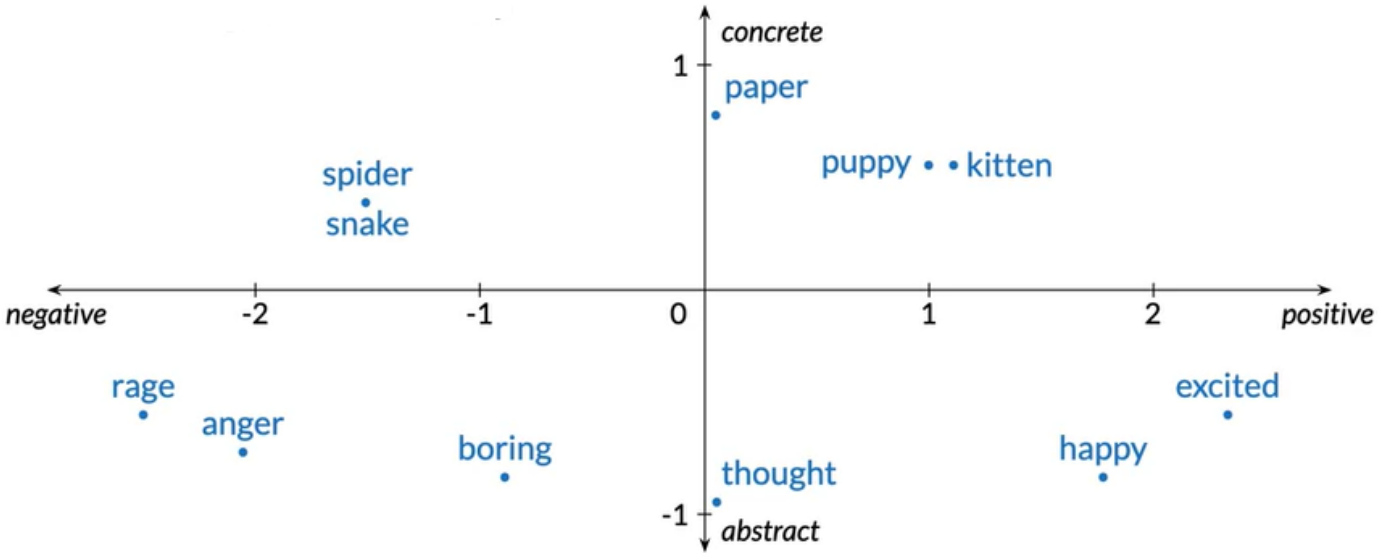
Now, you can say that's "happy" and "excited" are more similar to each other compared to with the word "paper"



Basic Word Representation – Word Embedding

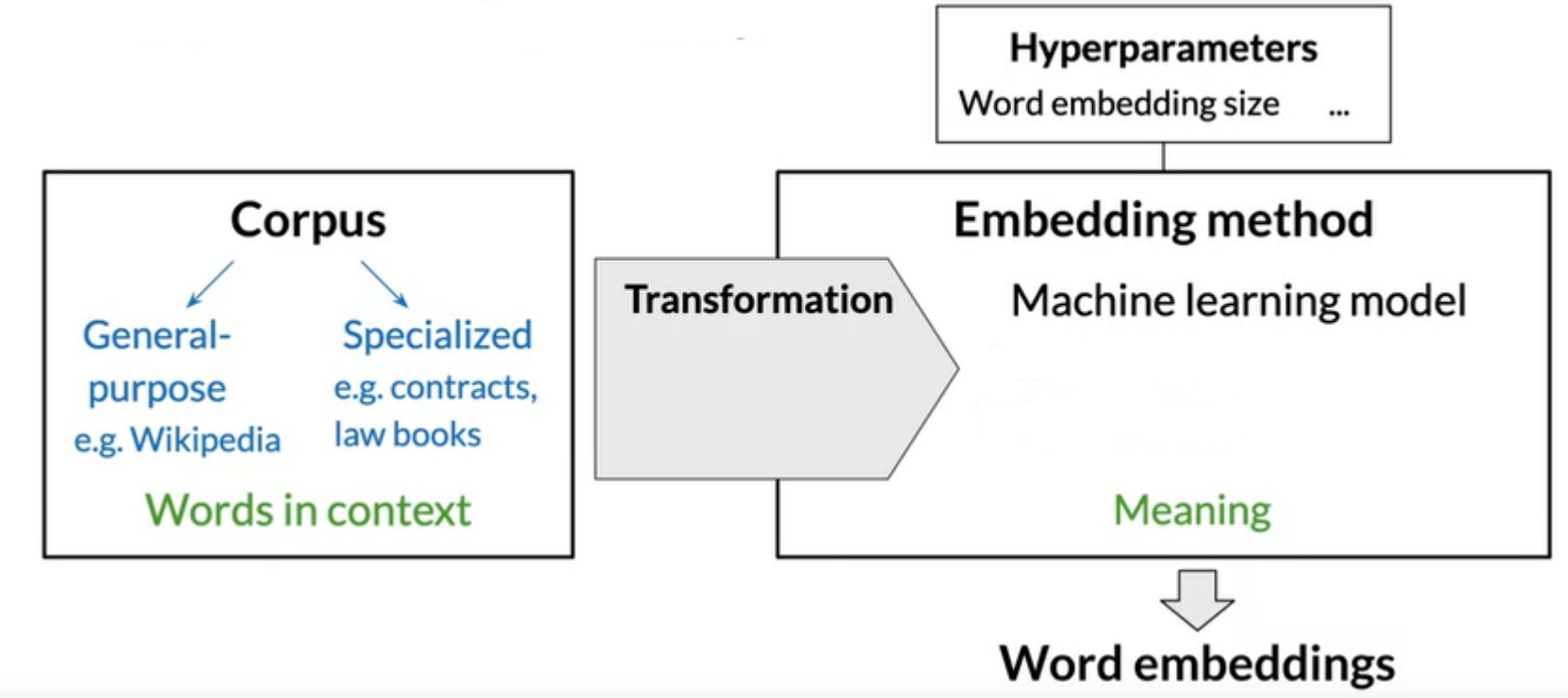
You can extend this by adding a vertical number line, words that are higher on this line are more concrete physical objects. Whereas words lower on this line are more abstract ideas.

What you've done here is to represent the vocabulary of words with a small vector of length 2.



Basic Word Representation – Word Embedding

- \+ Low dimension
- \+ Embed meaning
  - e.g. semantic distance

forest ≈ tree forest ≉ ticket

Basic Word Representation – Word Embedding Process